숙제 2: BST를 이용하여 두 파일의 유사도 검사

○ 문제 설명

- 두 파일(A와 B)의 이름을 입력받은 다음, 두 파일의 유사도를 출력한다.
- 먼저 "다섯 개의 연속된 단어" (shingle)를 key로 하고, 파일에서 해당 key의 발생 빈도수를 value로 하는 (key, value)쌍을 BST에 저장
  예) A의 내용: "it was the best of times it was the worst of times" 일 경우,
  첫 번째 key = "it was the best of"
  두 번째 key = "was the best of times"
  세 번째 key = "the best of times it"
  ...
  마지막 key = "was the worst of times"
- 파일 A와 B에 대해 BST를 각각 생성할 것
- 파일의 유사도는 ¦A∩B¦ / ¦A∪B¦ 로 정의.
  예) A의 BST에 {(x,3), (y,1)}이 저장되어 있고, B의 BST에 {(x,2), (z,2)}가 저장될 경우, A∩B = {(x,2)}이므로 ¦A∩B¦ = 2이고, A∪B = {(x,3), (y,1), (z,2)}이므로 ¦A∪B¦ = 6. 이 경우 A와 B의 유사도는 2/6.
- 토큰 추출을 위하여 다음과 같이 StringTokenizer 사용할 것.
  예: `StringTokenizer st = new StringTokenizer(str, " \t\n=;,<>()");`

○ 구현의 예

```
첫번째 파일 이름? h.java
두번째 파일 이름? t.java
파일 h.java의 Shingle의 수 = 123
파일 t.java의 Shingle의 수 = 109
두 파일에서 공통된 shingle의 수 = 43
h.java과 t.java의 유사도 = 0.2275132275132275
```

○ 과제 제출 방법: HW2.java **하나의 파일**만 제출

① public class HW2 (파일 내에 나머지 클래스들은 public이 아님)
② default package 사용
③ 주석은 모두 삭제
④ BST는 강의 노트에 나와 있는 클래스를 구현할 것
← **위의 조건들을 하나라도 만족하지 않는 과제물은 심사하지 않음!!**

○ 점수: 20점